September 1947

BIOMETRICS

Vol. 3 NO. 3

THE BIOMETRICS SECTION, AMERICAN STATISTICAL ASSOCIATION

THE DEVELOPMENT AND USE OF STATISTICAL PRACTICES IN HOSPITAL WORK*

PAUL M. DENSEN

Division of Medical Research Statistics, Department of Medicine and Surgery, Veterans Administration

As long ago as 1863 [23] Florence Nightingale[1] proposed the tabulation of "seven elements which are required to enable us to tabulate the results of hospital experience. The primary object of these tables was to obtain a uniform record of facts from which to deduce statistical results among which the following may be mentioned:

"1. The total *sick population*—i.e., the number of beds constantly occupied during the year by each disease for each age and sex.
"2. The *number of cases* of each age, sex, and disease submitted to (medical or surgical) treatment during the year.
"3. The *average duration* in days and parts of a day of each disease for each sex and age.
"4. The *mortality* from each disease for each sex and age.
"5. The annual proportion of *recoveries* to beds occupied and to cases treated for each age, sex, and disease."

REASONS FOR RECENT INTEREST

Today, nearly a century later, the barest beginning has been made along these lines and there are still many hospitals which would experience considerable difficulty in "deducing" all of the statistical results desired by Miss Nightingale. The seed had been planted but germination, or at any rate growth, has been slow indeed. From the stand-

* From the Department of Preventive Medicine and Public Health, Vanderbilt University School of Medicine, Nashville, Tennessee.

[1] For an account of Miss Nightingale as a statistician see reference [20].

point of identifying the elements which must exist in the environment for the healthy development of the statistical plant, it may be enlightening to inquire as to the reasons for this lag and for the recent upsurge of interest in this problem.

It would appear that a realization of the fact that the hospital "exercises a powerful influence on the pattern of medical and health care in a community" [7] is an essential. Steadily rising hospital costs per patient day coupled with rising admission rates [11] have focused attention on the need for more factual data to define the extent and nature of the problems which the hospitals are facing. Perhaps it is not too great an exaggeration to say that hospitals in some cases are like a health department which attempts to develop a sound program with no more knowledge of the health problems of the community than could be gained from the crude death rate. The problems which the hospitals face are population problems. Only by a much more intimate knowledge of the medical care problems of the community from the consumer's as well as the producer's standpoint is it likely that the hospital will be in a position to evaluate the future demands which may be made upon it and to discharge efficiently its functions as a community agency. Gradually the realization of the need for specific rates in studying these problems is having its influence in the development and use of statistical practices in hospital work.

Another force which has operated to produce a more widespread interest in hospital statistics than existed at the time Miss Nightingale wrote her "Notes" has been the increasingly quantitative nature of medicine. Writing in 1921, Pearl [25] has pointed out that "the entire history of medicine shows that there has been almost from the first an earnest desire and effort on the part of some of its leaders to develop quantitative modes of thought and methods of work. The large measure of progress which has been made in this direction is sufficiently evidenced by the number of items of diagnostic and clinical significance which are measured and recorded in quantitative terms." If this was true then, how much more so it must be now. If further

Entered as second-class matter, May 25, 1945, at the post office at Washington, D. C., under the Act of March 3, 1879. *Biometrics* is published four times a year—in March, June, September and December—by the American Statistical Association for its Biometrics Section. Editorial Office: 1603 K Street, N. W., Washington 6, D. C.

Membership dues in the American Statistical Association are $5.00 a year, of which $3.00 is for a year's subscription to the quarterly *Journal*, fifty cents is for a year's subscription to *The American Statistician*. Dues for Associate members of the Biometrics Section are $2.00 a year, of which $1.00 is for a year's subscription to *Biometrics*. Single copies of *Biometrics* are $1.00 each and annual subscriptions are $2.00. Subscriptions and applications for membership should be sent to the American Statistical Association, 1603 K Street, N. W., Washington 6, D. C.

evidence is needed, one has only to note the increasing number of medical schools with teaching hospitals attached which have incorporated courses in medical statistics in the curriculum [26]. In addition, therefore, to much greater production of quantitative data the medical man is being trained to think in terms of groups as well as in terms of the individual, all of which produces a favorable environment for the development of hospital statistics.

The lack of mechanical equipment for handling large masses of data probably has had a great deal to do with the tardy growth of statistics in the hospital field. It has really only been since the first World War that punch card equipment has come into widespread use. In the first of several papers from the newly organized statistical department of the Johns Hopkins Hospital, Raymond Pearl [25] described the general principles of the application of punch card techniques to hospital data. Since that time many other hospitals have applied these principles, some examples being the Mayo Clinic, the University Hospital at Ann Arbor, Michigan, the Columbia Presbyterian Hospital in New York City, the Massachusetts General Hospital, Vanderbilt University Hospital and many others. It should not be thought, however, that the punched card method is confined to large hospitals. Small hospitals can punch cards and arrange to have them tabulated elsewhere. Or, as we shall see later, several hospitals may engage in a cooperative arrangement.

BASIC ELEMENTS NEEDED FOR HOSPITAL STATISTICS

With the preceding discussion as a background, the basic elements needed for hospital statistics may be considered. The primary source of data is the hospital record of the patient. A hospital record's primary function is to help in the treatment of the patient. It is so intimately tied to the individual, however, that many of its research and study uses [13] may be overlooked because the emphasis in such work is on the mass of records and on the group of individuals they represent.

Even with respect to the group, hospital records have certain peculiarities not found in other types of records [8]. A hospital record has to take care of a tremendous variety of things. It is not directed toward obtaining information on one particular phase of the situation, such as a survey record usually is or an immunization record might be. Also the record is made out and used by a great many different people. Hence, it must be fairly simple and capable of being readily taught.

One of the greatest drawbacks to the statistical exploitation of the data in the hospital histories lies in the fact that emphasis in history taking is upon the positive findings. All too often negative findings are not recorded. On the other hand, the narrative form of the usual history makes it almost impossible for the physician to take the time to record all negative findings. This difficulty has been recognized for a long time and the remedy which has been proposed is to so design the form that a series of definite questions are asked for the items which are to form a part of every history [14, 25]. For other items whose appearance in the history depends upon the circumstances surrounding the individual case, blank space is left for the physician to record his observations in whatever form he pleases. Pearl [25] discusses at some length the objections which are likely to be raised to such a form and disposes of them in trenchant fashion. Dunn and Rockwood [15] designed a form of the type mentioned but so far as is known it has never been put to widespread use.

Why has it been so difficult to place into operation such forms which have been developed on sound principles? As has been pointed out above, it has only been very recently that the medical profession has been "statistically" conditioned to the point that the validity of the arguments advanced by Pearl and by Dunn and Rockwood may be appreciated by the medical man who writes the histories. There is also something of a problem in pedagogy here. Oftentimes, tabulations of records solely for the purpose of demonstrating their incompleteness have a very salutary effect upon future record keeping and result in a request *from* the medical man for a record designed along the principles exemplified in Dunn and Rockwood's forms.

The adoption of the unit record system has also aided in the development of the statistical uses of the records. This system provides that the first time a patient appears at the hospital, either as an in-patient or an out-patient he is assigned a number. This number remains unchanged no matter how often or over how long a period of time he continues to appear at the hospital and all observations which are made upon him are incorporated in one record which bears this number.

The significance of the unit medical record is well expressed by Kurtz [21]. To her "it is . . . the practical expression of a fundamental medical concept, that the individual—not some part of him or some episode in his history, but the whole individual—is the *unit* of medical practice and study." Until this concept was introduced into

the organization of medical records any attempt to bring together all the cases of a given condition or to engage in follow-up studies was almost a hopeless task, and statistical practice in relation to hospital records suffered accordingly. Most hospitals at the present time have adopted the unit record system, and new statistical uses are constantly being found for the data in the records.

Broadly speaking, the statistical uses to which the information in the hospital may be put can be categorized in terms of the universe of discourse to which they refer [10]. We may study, for example, the age distribution of individuals appearing in the syphilis clinic and diagnosed as having syphilis. But the generalization of such a study to all individuals appearing at the hospital or to the population at large would be totally unwarranted. The statistical universe here is limited to a particular kind of case appearing in the syphilis clinic. This is the type of universe from which most of the studies which are done by clinicians are drawn.

We may also be interested in the prevalence of various kinds of disease conditions among the different individuals who come to the hospital. We are then dealing with the hospital universe. Studies of data drawn from this universe have very definite uses [13], but such data will give us no idea of what the prevalence of these conditions is in the community. The difficulties which lie in such generalizations are well illustrated by Berkson [3] in his discussion of the limitations of the application of 4-fold table analysis to hospital data.

A third universe which must be recognized in relation to hospital statistics is the population from which the individuals coming to the hospital are drawn, or, to put it another way, "the population which the hospital is serving" [10]. For any one hospital, this is usually an extremely difficult population to define quantitatively. The task is somewhat simplified when all the hospitals in a community are considered.

CLASSIFICATION PROBLEMS

Attempts in the past to study samples drawn from these different universes have resulted in the recognition of certain technical statistical problems which needed to be solved before any great amount of headway could be made in the analysis of hospital data. Most of these are problems of classification.

Consider, for example, the clinician who wishes to work with a sample drawn from the "case" universe. He wishes to get all the

cases of a given kind together. It is important for such a study that his colleagues use the same term for indications of the same set of conditions. It is especially important if any attempt is to be made to combine the statistics of several hospitals. The experience of the Hospital Discharge Study [18] of New York City serves to illustrate the point. "The greatest difficulties were met in the terminology of cardiovascular conditions. Here more than anywhere else, symptoms, such as 'enlargement of the heart' or 'cardiac failure' replaced diagnoses. Vague notations—were so very frequent that there was no basis for further classification among the 25,758 cases reported with heart disease." This difficulty has long been recognized and the need for a standardized nomenclature of diseases was felt by Florence Nightingale when she sought the aid of Doctor Farr in the development of a morbidity list for use in hospitals. Massachusetts General Hospital and Bellevue Hospital in New York City developed their own nomenclatures to meet this need around the turn of the century and these nomenclatures were adopted by many other hospitals in the country. By 1928 a number of other nomenclatures were also in use and the need for a unification of these nomenclatures resulted in the National Conference on Nomenclature of Disease which brought out a Standard Classified Nomenclature of Disease. The publication of this Standard Nomenclature has since been taken over by the American Medical Association [22] and by 1935 the Nomenclature had been adopted by nearly 500 hospitals in the United States and Canada.

A nomenclature, however, is only the first step in the study of problems related to the "hospital" universe or to the population which the hospital serves. "The function of a nomenclature is to train the physician to use the clearest and most acceptable diagnostic terms to describe a particular clinical case." A classification of disease for statistical purposes is also needed to group the thousands of medical diagnoses so that they may be presented in meaningful tabular form and so that with the aid of an alphabetical index "a reasonably intelligent diagnosis coder may assign diagnostic statements to the various categories of the list as accurately as is possible from the stated causes of illness" [24]. In 1936 Berkson [1, 2] presented such a classification and a plan for meeting, by means of one punch card, the needs of the clinician who wishes to get all the cases of a particular diagnosis and, also, the statistical requirements of periodic statistical summaries of medical conditions. This scheme is an ingenious use of a punch card code consisting, for the diagnostic items, essentially of a main

number and a sub-number, the main number serving the statistical needs and the sub-number permitting the identification of cases of a particular diagnosis. The punch card is also designed to permit writing the diagnoses directly on the card and also the coding of the diagnoses.

The International List of Causes of Death served as the basis for Berkson's tabular outline for the classification of disease terms because it "has the widest current use for purposes of statistical enumeration." The Welfare Council of New York also evolved a Classified List of Diagnoses in connection with its Hospital Discharge Study and this has been reported by Jeter and Fraenkel [19] and by Fraenkel [16]. As a result of experience gained with these lists and of the experience of other groups a committee of consultants appointed by the Surgeon General of the United States Public Health Service evolved a more general "Diagnosis Code for Tabulating Morbidity Statistics" which appeared in 1940 [24].

At the Fifth International Conference for the Revision of the International List of Causes of Death held in Paris in 1938, it was recommended that "the United States Government continue its studies of the statistical treatment of joint causes of death". In compliance with this resolution the Secretary of State appointed a United States Committee on Joint Causes of Death and this committee "decided that before taking up the matter of joint causes it would be advantageous to consider classification of disease from the point of view of morbidity and mortality since the joint cause problem pertains to both types of statistics". Accordingly it appointed a subcommittee to draft a "classification that might be used for the statistical coding of both records of illness and causes of death". The labors of this group brought forth in April 1946 a Proposed Statistical Classification of Diseases, Injuries and Causes of Death [27] which was adopted by the U. S. Committee on Joint Causes of Death. After a number of further corrections the parent committee turned over the Proposed Classification to the Interim Commission of the World Health Organization in March 1947 at the request of the latter. The introduction to the Proposed Classification contains an excellent history of the development of classifications of disease. The Proposed Classification itself is perhaps as great a step forward as the original development of the International List of Causes of Death. It is to be hoped it will be as widely adopted.

With the development of classifications of disease, knowledge of

the extent and causes of illness in the population also began to grow. Most of this knowledge is the result of special surveys of one type or another except for such information as is furnished by the reporting of certain notifiable diseases. It has already been indicated in the opening paragraph of this review that the idea of combining the statistics of several hospitals to furnish morbidity data is not a new one. The universe to which such data apply, however, must be clearly recognized and the evolution of such statistics mirrors the constant striving to enlarge the universe to as much of the population at large as possible.

STUDIES OF HOSPITAL MORBIDITY

In 1913 Frederick L. Hoffman presented the statistics of the Johns Hopkins Hospital from 1892 to 1911 [17]. This monograph was an excellent demonstration of what could be done in studying the records of one hospital. The universe to which it applied, however, is the "hospital universe" only. In the same year Bolduan [4] proposed the adoption of a plan for collecting morbidity statistics from all the hospitals in a community utilizing a procedure similar to that by which data on death certificates were compiled. Essentially his idea was to have a "discharge certificate" sent to a central collecting agency each time a patient was discharged from the hospital. The central agency would then code and tabulate the data on the certificate. The greatest deterrent to the adoption of this scheme was the difficulties which were encountered in classifying the data in the different hospitals. However, this plan served as the basis for a study of a sample of 21,000 patients in six hospitals in New York City in 1923. Its principle was also adopted in the Hospital Discharge Study of the Welfare Council of New York City published in 1942 in which an analysis was made of over half a million patients discharged from the hospitals of the city. Insofar as this material covered most of the hospital facilities in the city, to that extent can rates based upon the population of the city be considered a reflection of the hospitalized illness which occurred in the city as a whole. It must be remembered, however, that this does not give a picture of total morbidity because only the more serious illnesses are hospitalized. As has been pointed out above, this study was especially valuable in the development of a morbidity classification for use in problems of this kind.

In 1939 Crosby [9] presented the results of a study of rural hospital morbidity. It was emphasized that the data applied to the "hos-

pital universe'' only, but this study did for a rural hospital what Hoffman's study did for the Johns Hopkins Hospital.

The studies mentioned up to this point have all started with a study of patients who had already been in the hospital. But in 1941 [11] and 1943 [12] Councell presented the results of a study which combines ''the use of survey and hospital data in that it starts with a survey group and proceeds to hospital sources of information for these persons''. It has been pointed out that the population which a hospital serves is difficult to define. Here the idea of taking a known population and finding out what hospital service it receives has been explored. The recent growth of hospital service plans, such as the Blue Cross, is making a wealth of material available for study of these problems. The reports of Van Dyk [27] and Colman [5, 6] show what can be done with these data. Here again, ''although these records pertain to a well-defined population, the rates are biased in that the persons studied are a selected group. Admission figures are higher than in the general population because, by the prepayment of fees, the economic barrier to hospitalization is considerably broken down.'' One still largely untapped source of morbidity statistics is the records of outpatients.

It has been shown how in recent years many of the obstacles to the development of hospital statistics have been broken down. Classifications of disease have been developed which permit more ready combination of the statistics of many hospitals, mechanical procedures for handling large masses of data have been perfected, and the medical man and medical environment have become more and more statistically minded and appreciative of the requirements of hospital statistics. Much still needs to be done, but hospital statistics are on the threshold of a tremendous development and ''hospital horizons are expanding'' [7]. It is to be hoped that the need for more trained personnel in this area will be recognized by those young people who are seeking fields in which they may pursue a promising career.

REFERENCES

[1.] Berkson, J., "A system of codification of medical diagnoses for application to punch cards, with a plan of operation," *American Journal of Public Health*, 26 :606–612, 1936.

[2.] Berkson, J., "Tabular outline for use in reporting hospital morbidity," *Proceedings, Staff Meeting, Mayo Clinic*, 11 :396–400, 1936.

[3.] Berkson, J., "Limitations of the application of the fourfold table analysis to hospital data," *Biometrics Bulletin*, 2 :47–53, 1946.

[4.] Bolduan, C. F., "Hospital morbidity statistics, a practical method of making them uniform and preparing them for analytic study," *New York Medical Journal*, 98 :643–646, 1913.

[5.] Colman, J. D., "Variations in experience classified by length of enrollment," *Proceedings, Mid-winter Conference Hospital Service Plan Executives,* Pittsburgh, Jan. 24–26, 1940, pp. 47–51.
[6.] Colman, J. D., "Age and sex variations in hospital utilization," Paper presented at *74th Annual Meeting of the American Public Health Association,* Cleveland, Nov. 11–14, 1946.
[7.] Corwin, E. H. L., *The American Hospital,* The Commonwealth Fund, New York, 1946.
[8.] Crosby, E. L., Unpublished lectures.
[9.] Crosby, E. L., "Rural hospital morbidity: statistical analysis of 13,014 consecutive hospital discharges in Otsego County, New York," *Rural Medicine,* 26–41, Charles C. Thomas, Balto., 1939.
[10.] Crosby, E. L., Discussion of reference 9 following all papers on rural morbidity in same book.
[11.] Councell, Clara E., "Hospital records as a source of morbidity statistics," *American Journal of Public Health,* 31:1044–1050, 1941.
[12.] Councell, Clara E., "Past hospital experience of a surviving population, Eastern Health District, Baltimore, 1926–1935," *American Journal of Hygiene,* 38:8–26, 1943.
[13.] Densen, P. M., "The administrative and research value of routine analyses of hospital statistics," Paper presented at *74th Annual Meeting of the American Public Health Association,* Cleveland, Nov. 11–14, 1946.
[14.] Dunn, H. L., "Status of statistical medicine," *Journal of the American Medical Association,* 1274, 1927.
[15.] Dunn, H. L. and Rockwood, R., "A record system suitable for both clinical and statistical medicine," *Archives of Internal Medicine,* 41:499–535, 1928.
[16.] Fraenkel, Marta, "Classification of hospital morbidity; experience in a study of hospital discharges in New York City," *Journal of the American Statistical Association,* 35:489–497, 1940.
[17.] Hoffman, F. L., *Statistical Experience Data of the Johns Hopkins Hospital, 1892–1911,* Johns Hopkins Press, Balto., 1913.
[18.] *Hospital Discharge Study,* Welfare Council of New York City, 1942.
[19.] Jeter, H. R. and Fraenkel, M., "Hospital morbidity reporting," *American Journal of Public Health,* 30, 1940.
[20.] Kopf, E. W., "Florence Nightingale as statistician," *Quarterly Publication of the American Statistical Association,* Dec. 1916.
[21.] Kurtz, D. L., *Unit Medical Records,* Columbia University Press, New York, 1943.
[22.] Logie, H. B., *Standard Classified Nomenclature of Disease,* American Medical Association, Chicago, 1937.
[23.] Nightingale, F., *Notes on Hospitals,* Third Edition. Longman, Green, Longman, Roberts and Green. London, 1863.
[24.] Parran, T. and Austin, W. L., "A diagnos . code for use in tabulating morbidity statistics," *U. S. Public Health Reports,* 55:1558–1575, 1940.
[25.] Pearl, R. E., "Modern methods in handling hospital statistics," *Johns Hopkins Hospital Bulletin,* 32:184, 1921.
[26.] Roberts, F. L., "Teaching of statistical analysis in medical schools," *Journal of the Association of American Medical Colleges,* Jan. 1941.
[27.] U. S. Committee on Joint Causes of Death, *Proposed Statistical Classification of Diseases, Injuries, and Causes of Death,* Wash., D. C., April 1946.
[28.] Van Dyk, F., "Experience as a guide in the development of group hospitalization," *Hospitals,* 12:79–83, 1938.

PROBABILITY TABLES FOR INDIVIDUAL COMPARISONS BY RANKING METHODS

FRANK WILCOXON

American Cyanamid Company

In previous papers [3] [4] a ranking method has been described for the rapid approximate determination of the significance of differences between two treatments, when the experimental data consists of unpaired replications, paired replications, or replications occurring in two or more groups. Brief tables were given which furnished the rank totals for the 0.05%, 0.02% and 0.01% level of significance. These tables were prepared by making use of certain properties of the partitions of numbers, but the method becomes impractical with larger numbers of replicates. A recent paper by Mann and Whitney [1] describes a similar test and gives probability tables covering the range from 3 to 8 replicates per treatment, including the case of unequal numbers of unpaired replicates under the two treatments. These authors tabulated the probabilities against the serial number of possible rank totals, 0–1–2 . . . . U, instead of the rank total itself. Their tables give probabilities for one tail of the distribution only.

UNPAIRED REPLICATES

Let N be the number of replicates, then $2N$ will be the number of rank scores from 1 to $2N$ to be assigned to the data. If the two treatments do not differ, then the total score for one of the treatments in repeated experiments will be distributed around the expected total $\overline{T}$ with variance equal to $N\overline{T}/6$. This may be shown in the following manner:

The variance of a single item from the rectangular population 1 to $2N$ is $(4N^2-1)/12$ [2]. The variance S^2 of a total of N items would be $N(4N^2-1)/12$, except for the fact that no rank number can occur more than once; in other words it is a drawing "without replacement." This reduces the variance by a factor $1-(N-1)/(2N-1)$, or $N/(2N-1)$. The corrected variance is $N^2(4N^2-1)/12(2N-1)$. If the corrected variance is divided by the expected total $\overline{T}$ which may be written $2N\ (2N+1)/4$ we obtain $N/6$ for the ratio $S^2/\overline{T}$, and hence $S^2 = N\overline{T}/6$.

The distribution of rank totals obtained by this procedure are sufficiently close to normal so that the totals corresponding to the 0.05%,

0.02%, and 0.01% levels may be closely approximated by means of the following formulae:

$$T_{.05} = \overline{T} - 1.960\ \sqrt{N\overline{T}/6}$$
$$T_{.02} = \overline{T} - 2.326\ \sqrt{N\overline{T}/6}$$
$$T_{.01} = \overline{T} - 2.576\ \sqrt{N\overline{T}/6}$$

In the case of the 0.001% level, however, the approximation is not sufficiently close to be useful.

PAIRED REPLICATES

In this case N rank numbers are assigned to the differences between pairs, and each rank number is given a sign corresponding to the sign of the difference. The expected total of one sign, $\overline{T}$, is $N(N+1)/4$, and the variance of the total is $2N\overline{T}/6$. The totals corresponding to the 0.05, 0.02, and 0.01% level of significance may be calculated as indicated previously in the case of unpaired replicates:

$$T_{.05} = \overline{T} - 1.960\ \sqrt{2N\overline{T}/6}$$
$$T_{.02} = \overline{T} - 2.326\ \sqrt{2N\overline{T}/6}$$
$$T_{.01} = \overline{T} - 2.576\ \sqrt{2N\overline{T}/6}$$

GROUPED DATA

Frequently the experimental results fall into two or more groups, with two or more replicates per group, as for example when comparisons are carried out at different concentrations, times, localities, or temperatures. If we consider one of the n groups, and assign rank numbers to the $2N$ results within a group in the manner described under unpaired replicates, the expected total, $\overline{T}$, for a treatment in one group is $2N(2N+1)/4$ and the variance is $N\overline{T}/6$. The expected total for n groups each with N replicates is $n\overline{T}$, and the variance of the total is $nN\overline{T}/6$ since the totals are additive and the variances also. The formulae for calculating the total corresponding to the 0.05%, 0.02%, 0.01% level are:

$$T_{.05} = n\overline{T} - 1.960\ \sqrt{nN\overline{T}/6}$$
$$T_{.02} = n\overline{T} - 2.326\ \sqrt{nN\overline{T}/6}$$
$$T_{.01} = n\overline{T} - 2.576\ \sqrt{nN\overline{T}/6}$$

where n is the number of groups, N the number of replicates per group, and T is the expected total for one group.

TABLE I

UNPAIRED REPLICATES

Probability (P) of chance occurrence of a rank total equal to or less than T with N replicates. T is given in body of table, to nearest whole number.

N	$P = 0.05$	$P = 0.02$	$P = 0.01$
5	18	16	15
6	27	24	23
7	37	34	32
8	49	46	43
9	63	59	56
10	79	74	71
11	97	91	87
12	116	110	105
13	137	130	125
14	160	152	147
15	185	176	170
16	212	202	196
17	241	230	223
18	271	259	252
19	303	291	282
20	338	324	315

TABLE II

PAIRED REPLICATES

Probability (P) of a chance occurrence of rank total of one sign, + or −, whichever is least, equal to or less than T. T is given in body of table, to nearest whole number. N is number of replicates.

N	$P = 0.05$	$P = 0.02$	$P = 0.01$
6	1	0	.....
7	2	1	.....
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	18	13	9
14	22	16	12
15	26	20	15
16	31	24	19
17	36	28	23
18	41	33	27
19	47	38	32
20	53	43	37

With unequal numbers of replicates in the different groups the expected totals and variances must be calculated separately and added together, to obtain the final expected total and its variance.

TABLE III

Grouped Data

Probability P of chance occurrence of a rank total equal to or less than T with N replicates in n groups:—

N	$n=2$		$n=3$		$n=4$		$n=5$		$n=6$		$n=7$	
	T	P	T	P	T	P	T	P	T	P	T	P
2	.....	0.01	9	0.01	13	0.01	17	0.01	22	0.01	26	0.01
	.....	0.02	10	0.02	14	0.02	18	0.02	23	0.02	27	0.02
	6	0.05	11	0.05	15	0.05	19	0.05	24	0.05	28	0.05
3	13	0.01	21	0.01	30	0.01	39	0.01	49	0.01	58	0.01
	14	0.02	22	0.02	31	0.02	41	0.02	50	0.02	59	0.02
	15	0.05	24	0.05	33	0.05	42	0.05	52	0.05	62	0.05
4	24	0.01	39	0.01	54	0.01	70	0.01	86	0.01	102	0.01
	25	0.02	40	0.02	56	0.02	72	0.02	88	0.02	105	0.02
	26	0.05	42	0.05	58	0.05	75	0.05	91	0.05	108	0.05
5	38	0.01	61	0.01	85	0.01	110	0.01	135	0.01	160	0.01
	39	0.02	63	0.02	88	0.02	112	0.02	138	0.02	163	0.02
	42	0.05	66	0.05	91	0.05	116	0.05	142	0.05	168	0.05
6	55	0.01	89	0.01	124	0.01	159	0.01	195	0.01	231	0.01
	57	0.02	92	0.02	127	0.02	162	0.02	198	0.02	235	0.02
	61	0.05	96	0.05	131	0.05	168	0.05	204	0.05	241	0.05
7	77	0.01	123	0.01	170	0.01	218	0.01	266	0.01	314	0.01
	79	0.02	126	0.02	174	0.02	222	0.02	270	0.02	319	0.02
	83	0.05	131	0.05	179	0.05	228	0.05	277	0.05	327	0.05

REFERENCES

[1] Mann, H. B. and Whitney, D. R., "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *Annals of Mathematical Statistics, 18*, pp. 50–60, 1947.

[2] Smith, James G. and Acheson, J. Duncan, *Sampling Statistics and Applications*, page 56, McGraw-Hill Book Company, Inc., New York, 1945.

[3] Wilcoxon, Frank, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin 1*, No. 6, pp 80–83, 1945.

[4] Wilcoxon, Frank, "Individual Comparisons of Grouped Data by Ranking Methods," *Journal of Economic Entomology*, 39(2) : 269, 1946.

TREATMENT OF THE FOURFOLD TABLE BY PARTIAL ASSOCIATION AND PARTIAL CORRELATION AS IT RELATES TO PUBLIC HEALTH PROBLEMS

HERBERT L. LOMBARD, M.D., AND CARL R. DOERING, M.D.

Massachusetts Department of Public Health and the Harvard School of Public Health

Many public health studies require treatment of plural variables and the bulk of data which is available falls into a dichotomous classification. The following are examples of studies which have required analysis by a technique applicable to alternative categories:

(a) An estimate of the risk of operative mortality for cancer patients was desired [1]. Obesity, malnutrition, old age, hypertension, cardiac history and length of operation were the variables studied. Many of these were not quantitatively measurable, and for consistency the few that were measurable were consolidated into alternative categories.

(b) In another study aimed to determine the amount of cancer knowledge in relation to age and economic factors, the age could have been measured quantitatively, but the economic condition could not [2].

(c) At present a study is being conducted in which approximately 80 variables dealing with heredity and environment as possible causative factors of cancer are being considered. Over three-quarters of these variables cannot be measured quantitatively and the solution of the many problems inherent in this study depends on the treatment of the 2×2 table.

If all variables in a study were independent, the examination of the data in the zero order of association would be sufficient. In the presence of data requiring the analysis of plural dependent variables, the two methods to be considered are partial association and partial correlation. The one is the effort to eliminate the fallacy of mixed classification by using partial universes; the other is the effort to eliminate the effect of certain variables in the whole universe, so as to be free of the fallacy of mixed classification.

With a satisfactory coefficient of correlation the method of partial correlation could be used. This would be of great advantage in those studies in which the data are not sufficiently numerous to enable subdivision of all the variables. If the data are numerous enough, partial association could be used. This is the association between x and y

TABLE I

EXAMPLE OF UNIT DEVIATE METHODOLOGY*

Newspapers

	Radio						No Radio					
	Solid Reading			No Solid Reading			Solid Reading			No Solid Reading		
	G†	P	T	G	P	T	G	P	T	G	P	T
Lectures	23	8	31	8	4	12	27	18	45	7	6	13
No Lectures	102	67	169	35	59	94	201	177	378	75	156	231
Total	125	75	200	43	63	106	228	195	423	82	162	244
	18.40 10.67 15.50			18.60 6.35 11.32			11.84 9.23 10.64			8.54 3.70 5.63		
$P_1 - P_2$	7.73 ± 5.29			12.25 ± 6.27			2.61 ± 3.01			4.84 ± 3.12		
$\frac{X}{\sigma}$	1.46			1.95			.87			1.55		

No Newspapers

	Radio						No Radio					
	Solid Reading			No Solid Reading			Solid Reading			No Solid Reading		
	G	P	T	G	P	T	G	P	T	G	P	T
Lectures	1	3	4	4	3	7	3	8	11	2	10	12
No Lectures	16	16	32	13	50	63	67	83	150	84	393	477
Total	17	19	36	17	53	70	70	91	161	86	403	489
	5.88 15.79 11.11			23.52 5.66 10.00			4.29 8.79 6.83			2.33 2.48 2.45		
$P_1 - P_2$	−9.91 ± 10.50			17.86 ± 8.36			−4.50 ± 4.01			−.15 ± 1.84		
$\frac{X}{\sigma}$	−.94			2.14			−1.13			−.08		

$$\Sigma\frac{X}{\sigma} = 5.82 \qquad \frac{\Sigma\frac{X}{\sigma}}{n} = .728 \qquad \frac{1}{\sqrt{8}} = .354 \qquad t = \pm\frac{.728}{.354} = \pm 2.06$$

* 1763 Records collected.
1729 Records had information on all variables.

† G = Good Score. P = Poor Score. T = Total.
n = all variables.

in sub-universes. The more variables to be considered, the greater will be the number of sub-universes to be studied. The elimination of the effect of one variable requires two sub-universes; two variables, four sub-universes; three variables, eight; four variables, sixteen, et cetera.

No additional computations are necessary when all partial universes show either direct or inverse association. When the values of the individual subdivisions are not significant, or where some are significant and others are not, it is necessary to use some method that will determine the significance of the subdivisions in the aggregate. A generally accepted method is to sum all the chi squares to form a single value and to look up the probability in the table of chi square with n the number of individual chi square values [3]. We have used this method in those instances in which all the subdivisions had either all positive or all negative associations. Unfortunately this rarely occurs if several variables are to be eliminated.

In recent public health studies [1, 2] the significance of the subdivisions in the aggregate has been determined by a slight modification of the method suggested by Professor E. B. Wilson of Harvard University. This consisted of obtaining $\frac{X}{\sigma}$ for each fourfold table comprising the subdivisions, averaging these unit deviates and determining significance, $\frac{M}{\sigma M}$. As the standard error of the normal curve is unity, $\frac{1}{\sqrt{N}}$ was used for the standard error of the mean. If the $\frac{M}{\sigma M}$ equalled or exceeded 2.6, significance was assumed. While $\frac{X}{\sigma}$ was usually the computed $\frac{P_1 - P_2}{\sigma(P_1 - P_2)}$, in those studies where the figures in the 2×2 tables were small, chi squares were computed using Yates' correction, the square roots of the various chi squares were extracted, and the sign adopted as if $\frac{P_1 - P_2}{\sigma(P_1 - P_2)}$ had been computed. By this method the benefit of Yates' correction for small numbers was obtained and the value of $\frac{X}{\sigma}$ was slightly smaller.

This is an excellent method if a considerable amount of data is available and if it is desired to eliminate the effect of several variables. The population in each subdivision should not be too small, and there should be enough subdivisions to give stability to the mean. In those

cases in which the size of the sample will not warrant subdividing the data into as many orders as desirable, or where only one or two variables are to be eliminated, some other method of analysis seems necessary.

Various coefficients of association for the 2×2 table have been used, but there are only two correlation coefficients: the tetrachoric r and the product-moment correlation coefficient described by Yule and

TABLE II

EXAMPLE OF THE PRODUCT-MOMENT CORRELATION COEFFICIENT FOR A 2×2 TABLE*

	Good Score	Poor Score	
Lectures	75	62	137
No Lectures	605	1020	1625
	680	1082	1762*

$$r_{12} = \frac{N\delta}{\sqrt{(A)(a)(B)(\beta)}}$$

$$\delta = (AB) - \frac{(A)(B)}{N}$$

$$\delta = 75 - \frac{680 \times 137}{1762} = 22.2$$

$$r_{12} = \frac{22.2 \times 1762}{\sqrt{680 \times 1082 \times 1625 \times 137}} = .096$$

The other zero corelations are:

Good score and radio	.118
Good score and solid reading	.296
Good score and newspapers	.246
Lectures and radio	.110
Lectures and solid reading	.117
Lectures and newspapers	.106
Radio and solid reading	.109
Solid reading and newspapers	.380
Radio and newspapers	.202

$r_{12}.345 = .0502$

$z = .05028$

$$\sigma_z = \frac{1}{\sqrt{1729 - 3 - 3}} = .0241$$

$$\frac{z}{\sigma_z} = 2.09$$

* 1763 records collected.
1729 records had information on all variables.

Kendall [3]. In using the tetrachoric method of correlation, the underlying assumption that the variables under consideration are really continuous and normal in distribution must be borne in mind, as well as the fact that the data for each trait have been forced into two alternative categories.

At the beginning of the heredity and environment study the tetrachoric coefficient computed from the diagrams prepared by Thurstone and his associates [4] was used in the hope that the important variables had a normal distribution. It became evident that this was not the case and the coefficient was therefore abandoned. The product-moment correlation coefficient for the 2×2 table was tried next and partial correlations computed. This was wholly empirical as it was

not known whether the coefficient could be used in partial correlations. After the partials were computed they were transformed into Fisher's z [5], and the results reduced to unit deviates by dividing each z by its standard error.

Numerous comparisons between partial correlation and partial association were made. The results of a few of these comparisons are shown in Table III.

TABLE III

COMPARISON OF RESULTS USING PARTIAL ASSOCIATION AND PARTIAL CORRELATION OF VARIOUS DATA

Variables	Number of Variables Eliminated	$\Sigma \frac{\frac{x}{\sigma}}{\frac{M}{1}}{\sqrt{N}}$	$\frac{z}{\sigma_z}$
Study A—436 records			
Cancer of cervix—use of spices	2	3.96	3.92
Cancer of cervix—irregular eating	2	4.59	4.54
Cancer of cervix—little hair	2	5.58	6.00
Cancer of cervix—heavy menstruation	2	3.17	3.67
Study B—493 records			
Cancer of breast—little hair	2	4.10	5.06
Cancer of breast—infrequent bathing	2	5.87	5.90
Study C—1729 records			
Knowledge of cancer—radio addresses	3	2.22	2.60
Knowledge of cancer—magazines & books	3	5.76	9.40
Knowledge of cancer—lectures	3	2.06	2.09
Knowledge of cancer—newspapers	3	5.55	5.67
Study D—2091 records			
Operative mortality—long operation	3	9.90	12.00
Operative mortality—heart disease	3	2.39	3.30
Study E—10,092 records			
Chronic disease—laxatives	5	15.70	16.10

On the whole, agreement between results of the two methods seems satisfactory. Some of the comparisons showed considerable difference, but one would expect a few differences, some perhaps serious, due to chance alone. It appears that we can deal with the problem of plural variables by either method, and the results, considering the limitations of the methods, are reasonably consistent. We are satisfied when both methods show consistency. When the results differ as to significance, we feel less certain but would be inclined to favor partial correlation if n is small or if only one or two variables are being eliminated, and

to prefer the partial association technique if several variables are being eliminated and the size of the sample is sufficiently large.

CONCLUSIONS

(1) Public health studies frequently require the use of the fourfold table in the analysis of data containing plural variables.

(2) The tetrachoric r has only limited applicability since in most studies some of the variables do not have a normal distribution.

(3) The additive property of x^2 can be used if all subdivisions have either positive or negative association.

(4) Partial association is a satisfactory method of analysis if the elimination of several variables is desired and the number of cases is sufficient for adequate populations in the sub-universes.

(5) Partial correlation, using the product-moment correlation coefficient for the 2×2 table, furnishes results sufficiently consistent with those of partial association as to warrant the opinion that it is a satisfactory method.

(6) The choice of the method to be used depends upon the data available.

REFERENCES

[1] Johnson, Allen S., and Lombard, Herbert L., "The Estimation of Operative Risk in Patients with Cancer," *New England Journal of Medicine*, 224:759–762, 1941.

[2] Potter, Evelyn A., "The New Bedford Survey," *Bulletin of the American Cancer Society, Inc.*, 27:7 (July) 1945.

[3] Yule, G. Udny, and Kendall, M. G., *An Introduction to the Theory of Statistics*. Charles Griffin and Company, Limited, London, 1937, pp. 426–427 and pp. 252–253.

[4] Chesire, Leone, Saffir, Milton, and Thurstone, L. I., *Computing Diagrams for the Tetrachoric Correlation Coefficient*, The University of Chicago Bookstore, 1933.

[5] Fisher, R. A., *Statistical Methods for Research Workers*, Oliver and Boyd, London, 1946, p. 210.

TIME-SPECIFIC LIFE TABLES CONTRASTED WITH OBSERVED SURVIVORSHIP*

MARGARET MERRELL

School of Hygiene, The Johns Hopkins University

In obtaining life tables for any group of individuals, biological or physical, there are two basically different types of observation which are commonly used. One of these consists in the classification of individuals at a fixed time into separate age groups and then in the recording of deaths for all the various ages simultaneously. The life table derived from this evidence is a time-specific or static life table because the deaths for all the ages are recorded at the same time. The second type of observations consists in following a group of individuals born at approximately the same time, from birth to the death of the longest-lived member of the group. In this case the survivorship curve is determined by direct observation. It has been called a generation or cohort or fluent life table [1]. The life tables resulting from these two different types of evidence on longevity will, except under unusual circumstances, be different in form, and in any case will be quite different in meaning.

To consider an example of the two life tables, suppose we wish to study the present day survival of Ford automobiles. We may ask two different questions: what is the longevity of Fords at the current rates of survival, or what is the survival of the current model of Ford? That is, in one case we ask about current rates of cars existing at the *present time* and in the other about the *present model* and its future rates. To answer the first question we might get the date of manufacture (that is the age) of all the Fords registered on January 1, 1947, and during the following year get, by date of manufacture, the number of these cars irrevocably eliminated from service, that is, the deaths. Death in this case would be due to accidents, old age, or perhaps disease and malformations. From the age-specific death rates we could construct a survivorship curve for Fords for the year 1947. It would give us the way in which a group of Fords would survive if the 1947 rates pertained throughout their careers. The other life table could be obtained by following the 1947 models throughout their future existence to determine the number finally retired from service up to successive dates. This would give us the survivorship curve for the 1947 models, by direct subtraction. These two survivorship curves

* Paper No. 233 from the Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University.

would be quite different and in this familiar problem everyone will think of a number of reasons why this is so. The cars registered on January 1, 1947 will be residues of different lots of cars, each remnant having its own peculiar ability to survive 1947 conditions, depending on age, basic quality, damage due to past experience, and nature of repair. The common element for all of them is that their experience is being studied for the year 1947. On the other hand the dynamic life table for the 1947 models will have the experience of successive years of time but the survivorship curve will be that of cars manufactured in the same year, and therefore will have a certain homogeneity which the fixed-time life table does not possess.

All of the factors that I have mentioned as contributing to the difference in the two life tables for Fords have their direct counterpart in biological life tables. Inherent capacity, environmental hazards, methods of preventing and repairing damage, are continually shifting to a greater or less degree in the biological as well as in the physical world. The individuals existing at any time are survivors out of different lots of individuals and a single survivorship curve determined by putting together their experience in a given period will be different from the survivorship of a given lot of individuals followed through their lives. Thus for biological species also, the static and the dynamic life tables will be different in both form and meaning.

Yet frequently the choice of life table in this field is made solely on the issue of convenience. For long-lived forms, like man, it is not convenient to follow a group of individuals from birth to death, so for this reason alone we sometimes employ the static life table. For short-lived forms, like the beetle or fruit-fly, it is often more convenient to start with a given brood and follow it through life than to keep track of the ages and deaths among different broods. The question of convenience has sometimes led to the error of comparing a specified-time life table for man with a generation life table for some other biological form. The scientific question at issue is the important point in these problems and that life table should be set up which is relevant to the scientific question.

If we turn to human life tables we can see the effect of determining survivorship on a fixed-time basis as compared with a generation basis.

Figure 1 gives the survivorship of males in Massachusetts for fixed times approximately 10 years apart from 1890 to 1940. The 1890 life table, for example, shows how a hypothetical group would survive if it died off according to the 1890 age specific rates. We can see from this graph the improvement in the rates from 1890 to 1940.

But the graph also shows the other type of life table in giving the survivorship for males born in 1890. This was constructed in the following way. The people born in 1890 were infants in that year, 1 year olds in 1891, 2 year olds in 1892, etc. The rates of dying at these ages in these calendar years are therefore the rates pertaining to this group[1]

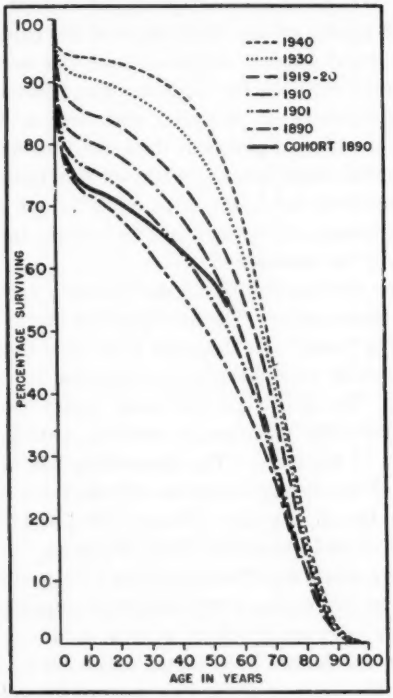

Figure 1. *Percentage surviving from birth to successive ages, Massachusetts males.*

as they pass through life. It is possible in terms of these rates to construct a survivorship curve for the 1890 cohort from birth to their present age of 56 years. I have carried this curve only to 50 years since the last life table I was able to construct on a known population distribution was for the census year of 1940. It is seen that the people born in 1890 had a survivorship curve unlike that of any of the years

[1] Although migration effects keep the composition of this group from remaining exactly the same in the sense of a laboratory experiment or an ideal follow-up, in a practical sense the same group is being followed.

through which they lived. Conversely the survivorship based on the rates of a given calendar year would not be the survivorship of any group followed throughout life.

Which of these curves do we want? If we want to compare the risks of dying in different calendar years we probably want the fixed-time life table. Whatever the multiplicity of factors in this year and in the past which produced the death rates of the calendar year, these rates were the actual current situation and the survivorship curve based on them helps us to see the implications of those rates. Certain aspects of the improvement in mortality with time, and the comparison of the mortality of different places at the same time are well shown by such curves. On the other hand, for certain prognosis problems, such as have been considered by Lotka [2] on the future age structure of our population, chances of orphanhood, and so on, the dynamic problem must certainly be considered.

If we consider the description of survivorship for other biological forms, the same issues arise. The question has been raised of getting the life table for a forest. In this case it is clear that the two forms of life table need to be very sharply distinguished since they would be widely different. The fixed-time life table would show the effect on survivorship of a current program of selective cutting, along with the natural mortality of the trees. The generation life table would show the survivorship of the same generation of trees from seedlings to final death influenced by all the time changes in forest preservation or destruction which would accompany their life span; this would give us the information we would need in evaluating a large scale reforestation program. The two life tables would therefore supplement each other in providing knowledge on which to plan a sound program.

If we consider the shorter-lived forms which are studied frequently in the laboratory, the contrast in the two types of life table is also great. The very fact of short life means that brief fluctuations in the environmental conditions may be very profound in their effect, and chance or deliberate alterations in environment will be different if they pertain to a single generation at some stage in its life span or if they affect a population composed of all ages.

I have been discussing a life table primarily in terms of a survivorship curve. I should like now to turn to a consideration of the age specific rates which are behind the survivorship curve and see how time is involved here. We study these age specific rates for their scientific meaning but the biological interpretation of the risk for a

given calendar year is very dependent on how the rates for successive generations have been changing.

I want to illustrate this by a study on tuberculosis, made by Dr. Wade Frost [3].

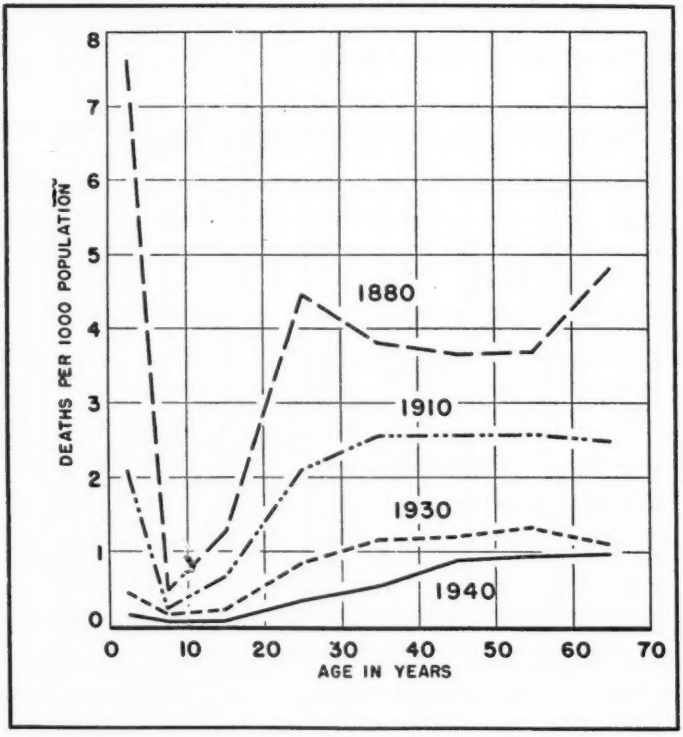

Figure 2. *Age-specific death rates from tuberculosis for different calendar years, Massachusetts males.*

Figure 2 shows the age-specific rates for males of Massachusetts for deaths from all forms of tuberculosis at certain fixed periods of time from 1880 to 1940. It had been noted by students of the subject that this curve had undergone a drastic change in shape. Back in 1880 the peak of mortality was in the 20's and at successive periods of time the peak became later and later in age until in 1940 it was in the 60's. Prior to Dr. Frost's work on this, the interpretation had

generally been that tuberculosis was changing in its relation to man, that at earlier times it was most devastating to young adults, but in recent years, either the disease or the resistance of people had changed so that the effects of the disease had become more serious later in life. In other words, tuberculosis had become an old age disease.

In order to test this explanation Dr. Frost examined the death rates for successive cohorts of people.

TABLE I

AGE SPECIFIC DEATH RATES PER 100,000 FROM TUBERCULOSIS FOR MASSACHUSETTS MALES, 1880 TO 1940, WITH RATES FOR COHORT OF 1880 INDICATED

Age	1880	1890	1900	1910	1920	1930	1940
0 - 4	760	578	309	209	103	41	11
5 - 9	43	49	31	21	24	11	2
10 - 19	126	115	90	63	49	21	4
20 - 29	444	361	288	207	149	81	35
30 - 39	378	368	296	253	164	115	51
40 - 49	364	336	253	253	175	118	86
50 - 59	366	325	267	252	171	127	92
60 - 69	475	346	304	246	172	95	109
70 +	672	396	343	163	127	95	79

The columns given in Table I are the age-specific death rates in successive time periods which we have just seen. If now we consider the people who were under 10 in 1880 as one group, (called the 1880 cohort), their death rates are in the 1880 column; they were 10 to 20 in 1890, and their death rate is in the 1890 column; 20 to 30 in 1900, and so on. The diagonal stepped lines indicate the death rates for this group as they pass through life. Similarly the 1890 cohort may be followed through to age 50–59; the 1900 cohort to age 40–49, and so on.

Now if we look at these cohort curves shown in Figure 3 we find that they all have their peak in the 20's, that they are all approximately the same shape, that there is no evidence at all that tuberculosis is becoming an old age disease. Each successive generation has its

greatest toll taken in the 20's. What then is the explanation for the apparent shift in the risks seen in Figure 2? It lies in the fact that in successive generations the rates have come down at all ages, the

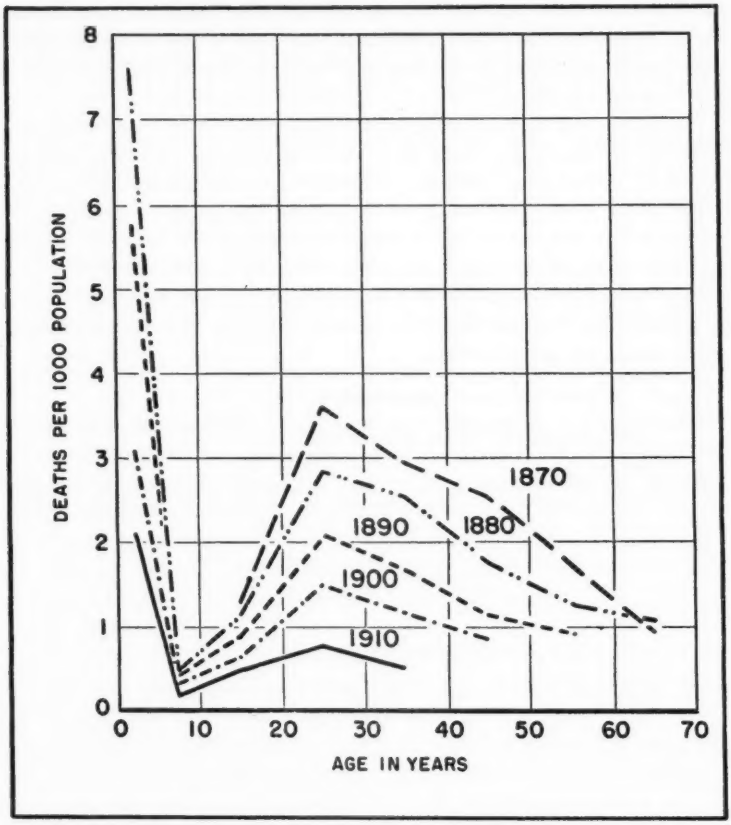

Figure 3. *Age-specific death rates from tuberculosis for successive 10-year cohorts, Massachusetts males.*

form of curve remaining the same. When we get the rates at a particular time, say 1940, we cut across the rates for all the various cohorts, at different points on the age scale. We have the 30-year-old rate from the 1910 cohort, the 40-year-old rate from the 1900 cohort, etc. The older rates pertain to earlier cohorts when *all* the rates were higher. Thus our 1940 curve does not represent the way tuberculosis

is affecting any single group of people in passing through their lives, but rather shows that the old people have come through such heavy exposure to the disease in their youth that there is more tuberculosis among them than among even our 20-year-olds today. It would not do to anticipate that our present 20-year-olds would show a higher rate in old age than they have now. Tuberculosis is still primarily a young person's disease.

It is the fact that age changes *in a person* are perfectly correlated with time changes that leads us to think of any age difference as representing a time flow. But age differences at a fixed time cannot be so interpreted. The very form in which the life table is put, that of survivorship, tempts us still further to think of the curve as giving us a flow of people through their lives. We must therefore distinguish clearly between the case where it does really represent the survivorship of a group and the case where it represents in very picturesque form, a static set of age-specific rates.

REFERENCES

[1] Dublin, Louis I, and Mortimer Spiegelman, "Current versus generation life tables," *Human Biology*, vol. 13, pp. 439–459, 1941.
[2] Lotka, A. J., "The structure of a growing population," *Human Biology*, vol. 3, pp. 459–493, 1931.
[3] Frost, W. H., "The age selection of mortality from tuberculosis in successive decades," *American Journal of Hygiene*, vol. 30, no. 3, sec. A, pp. 91–96, 1939.

PROFESSOR SNEDECOR RETIRES

G. W. Snedecor retired July 1 as head of the Statistical Laboratory, Ames, Iowa. He will remain associated with the Laboratory as research professor during the school year, continuing his teaching and consulting activities. His retirement comes 14 years after the Laboratory was founded in 1933. Professor Snedecor early took the lead and has done more than anyone else in this country towards promoting the application of modern statistical methods to biological research. He has a real understanding of the practical aspects in experimentation and insists on the use of common sense along with the statistical tools. We wish to take this opportunity to express our sincere personal appreciation for his patient instruction, wise counsel, and friendship. His influence holds a prominent place in the minds and purposes of many of us.

QUERIES

(51)

QUERY: In connection with certain investigations into railroad problems, it was necessary recently to determine whether the regression of certain series was linear or not. Fisher's summation method was used. Is it in order to use this method to determine if the regression is linear, and also is the test given conclusive?

Kindly state also if F is significant after compiling the second stage, it is still necessary to continue to calculate the third and fourth stages, etc. In other words, is it possible that if F is insignificant after the second stage for it again to become significant subsequently, which will, of course, mean that while the fit of a horizontal straight line is better than that of the fit given by a quadratic equation, yet the fit of an equation of higher degree may be closer than that of the quadratic.

ANSWER: My understanding is that you have values of Y corresponding to a set of values of X spaced at equal intervals, a single value of Y for each X. This may be a time-series. The summation method results in a set of additive mean squares, each associated with the coefficient of a polynomial. Some judgment as to the nature of the trend is arrived at by studying the graph and by relating it to the relative sizes of the successive mean squares.

Tests of significance of the orthogonal coefficients depend on the estimation of σ^2, the random component of variance: this σ^2 is the variance of a residual of Y, assumed normally distributed for each X and the same for all values of X; that is, it is independent of both X and Y. If my guess is correct, you have no very good method for estimating σ^2.

Lacking this, one method of procedure is to assume that some polynomial, of fifth degree for example, will follow the population regression with sufficient fidelity to warrant the use of the model. On this assumption the analysis of variances of n values of Y will take this form:

Source of Variation	Degrees of Freedom
Deviations from mean	1
Linear	1
Quadratic	1
Cubic	1
Quartic	1
Quintic	1
Remainder	$n-6$

Under the assumptions made, the remaining mean square with $(n-6)$ degrees of freedom is an estimate of σ^2, and may be used to test each orthogonal regression coefficient by use of t or F. Information from these tests should be used to supplement that gained from the graph.

You may decide that the linear component of the regression is predominant, the remaining curved components being negligible. In this sense, it may be said that you have "determined" (I'm a bit leery of that word) that the regression is linear.

To answer your last question, let us assume that you have found the last three orthogonal coefficients to be negligible, and have adopted the quadratic to represent the regression. In this case, the mean square associated with the coefficient of the linear polynomial might be large or small depending on the shape of that portion of the parabola lying within the bounds of your data. Here the situation could arise that the mean square for the second stage of fitting be small and non-significant, yet be followed by a large mean square in the third stage.

I hesitate to say that any of these tests are "conclusive" because I am not sure what you mean by the word. Any test involves a statement about probability and in a sense such statements can never be conclusive. On the other hand, if the mathematical model describes the population, then the statements about probability are exact; if these statements lead you to a conclusion, then you might refer to the test as conclusive. However, rational considerations are the final criteria, not tests of significance.

GEORGE W. SNEDECOR

(52)

QUERY: Upon reading the article by W. D. Baten and G. M. Trout entitled "A Critical Study of the Summation-of-difference-in-rank Method of Determining Proficiency in Judging Dairy Products" (*Biometrics Bulletin,* August 1946), I was impressed by the lack of symmetry in the distribution of the rank correlation coefficient as given in Table 3. It would appear from the form of the function that it would be symmetric. Upon referring to "The Advanced Theory of Statistics" by M. G. Kendall, page 396, I found a distribution Σd^2 for values of n running from 1 to 8. A linear transformation, $1-\frac{\Sigma d^2}{N(N^2-1)}$, upon the distribution for $n=7$ when grouped as was done in the *Biometrics* article gives the following result:

	Baten and Trout	M. G. Kendall
−1.1000 to −0.875	86	31
−0.625	266	319
−0.375	644	647
−0.125	976	975
+0.125	1096	1096
+0.375	985	975
+0.625	638	647
+0.875	321	319
+1.000	28	31

I should like to inquire if something else was intended by Baten and Trout.

I am somewhat surprised to find that both very high and very low values are considered significant with respect to judging dairy products.

ANSWER: No, nothing different was intended. The published table should have been as pointed out. Thanks for calling my attention to it. I do not know why it was published wrong. I take all the blame.

W. D. BATEN

(53)

QUERY: I have a 3×3 Latin square experiment in which, for the past 5 years, I have been comparing the effect of 3 kinds of manure on cultivated blueberries. I have proved the main point of this experiment, which is that manure is not sure death to cultivated blueberries, as was reported a good many years ago. The yields from the horse manure and cow manure plots are within a few quarts of each other so that it is obvious that there is no difference between these two. The yields from the poultry manure plots have been considerably below the other two. The question, then, is: Is poultry manure significantly inferior?

Since the blueberry is a perennial plant, how should the data be treated? Should the 5 years be added together and treated as a single year, or should they be kept separate?

Since several of the bushes on 2 of the plots were infected with a virus disease, how can appropriate comparisons be made?

ANSWER: In answer to your main question, the total yields for the 5 years should be used in testing the significance of differences among treatments, the structure of the analysis of variance being,

Rows	2
Columns	2
Treatments	2
Error	2

You are doubtless aware that, with so few degrees of freedom for error, you can expect little information beyond the fact that you mention—the plants were not killed. From this experiment, you don't even know whether the manures had any effect at all; that is, whether they changed yield from that of untreated plots or of plots fertilized with commercial fertilizer.

"Is poultry manure significantly inferior?" is a moot question which has been discussed before in this column (Vol. 1, page 26 and Vol. 2, page 16). From some research just completed by Dr. David B. Duncan, it seems that, in your particular case (two means at one end of the range with one at the other), it is appropriate to apply the t-test to the difference between the largest and smallest means.

In regard to the plants affected by disease, if it may be assumed that the other plants in the same plots were unaffected either by the disease or by lesser competition, it will be sufficiently accurate to use the average plot yields per plant in calculating the analysis of variance. With so few degrees of freedom for error, you are in a very vulnerable position; hence, any great refinements in statistical methods are superfluous.

GEORGE W. SNEDECOR

CORRECTION FOR QUERY (46), *Biometrics,* Vol. 3, No. 2, June, 1947: Page 95, line 9, omitting the word "not", should read:

"The method of estimating the missing value by minimizing SSE does give an unbiased estimate of σ^2."

NEWS AND NOTES

Our Biometrics Section chairman, D. B. DE LURY, has left Virginia Polytechnic Institute at Blacksburg to return to Canada. Cooler with no duties?! He is now with the Ontario Research Foundation, 43 Queens Park, Toronto, Canada. . . . The Ontario Research Foundation was established by Provincial Act in 1928 and is an endowed institution, half of its endowment being contributed by the Ontario Government and half by industry. A. E. R. WESTMAN, Director of Chemical Research, describes the work of the foundation briefly, "It carries on scientific research for industry, and to some extent for agriculture, forestry and fisheries in the fields of chemistry, metallurgy, biochemistry, textiles, parasitology, climatology and soils. During the war we were asked to assist in the control of manufacture of tank shoes." This work is described in an article by A. E. R. Westman and R. W. S. FREEMAN, "Statistical Control of the Manufacture of Manganese Steel Tank Shoes," *Canadian Metals and Metallurgical Industries*, June 1945. Mr. Westman says that there has been an increasing demand for help with statistical methods both within the Foundation and from other institutions and government departments. Call upon D. B. DeLury. . . . Reconstruction has been in progress in the University of Cambridge, England, since the return of the statisticians, who were all away on war service, the fort meantime being held by J. O. IRWIN who was in residence there. JOHN WISHART, reader in statistics, has resumed lecturing to mathematicians and to biologists. In the former sphere he is assisted by M. S. BARTLETT and H. E. DANIELS, lecturers in the Faculty of Mathematics. A Diploma in Mathematical Statistics has been established from 1 October 1947. This is a one-year course open to graduates with a reasonable standard in mathematics, who will study in the Statistical Laboratory (for which a temporary building will be constructed) the theory of statistics and its application to a selected one out of a wide number of fields ranging (alphabetically) from Agriculture to Psychology. In the biological sphere the University is appointing a Committee to consider what developments should be undertaken. In the Faculty of Economics and Politics, C. F. CARTER is lecturer in Statistics; and a Department of Applied Economics has started functioning under the Directorship of RICHARD STONE. Statistical problems in the Psychological Laboratory are looked after by E. G. CHAMBERS, and similar problems in relation to their particular sphere are beginning to engage the atten-

tion of the Department of Medicine. . . . RAE H. HARRIS, cereal technologist, Agricultural Experiment Station, Fargo, North Dakota—welcome to our ranks! He states, "We are using statistical methods in evaluating our departmental data, but I still feel that it is a difficult problem for the research worker who is not a professional statistician to select the proper techniques for securing the largest amount of information from his data." . . . It is possible we are seeking new subscribers to *Biometrics*. LEROY POWERS, Horticultural Field Station, Cheyenne, Wyoming, says, "I am sure that there is a real need for a journal of statistical methodology, and therefore, today am sending in my application for membership." During the last few months several letters have been received commenting on the kind of articles you, the readers, want. These comments are helpful and an effort will be made to secure such articles. . . . HAROLD F. DORN, Federal Security Agency, United States Public Health Service, Washington, D. C., gave some excellent suggestions. . . . P. A. MINGES, Extension Specialist in Truck Crops, College of Agriculture, Davis, California, seems to be falling in line with the rest of the people in California. He likes the West! However, he does falter maybe—"I do not believe that we raise quite as good strawberries as we did in Iowa, but of course one reason is that I am not working on that crop. My work deals primarily with vegetables." Let's hope no difficulties are introduced by publishing disloyalty. Come see what can be done in the South. . . . It has been observed that the acting editor published quotations which can have a variety of meanings when isolated from the text. Anyhow, it is some comfort to find that others who have visited in Honolulu understand. . . . CHAUNCEY D. LEAKE, Vice President, The University of Texas, Medical Branch, Galveston, enjoyed Hawaii a year ago in the spring. He writes, "We are very interested in the development of statistical analysis in our institution, and J. ALLEN SCOTT, Professor of Preventive Medicine and Statistician for John Sealy Hospital, is in charge of this effort for us." . . . JOHN C. ANDERSON, Assistant Research Specialist in Farm Crops, New Jersey Agricultural Experiment Station, New Brunswick, writes, "Professor Snedecor's answers to various questions, I find quite enlightening." . . . R. E. PATTERSON is now Assistant Director, Texas Agricultural Experiment Station, College Station. Does that mean he will have added interest in statistical methods? . . . JOSEF BROZEK, Laboratory of Physiological Hygiene, University of Minnesota, Minneapolis, writes, "One of the functions of Biometrics should be to spread biometrical methods to all fields in

which quantitative data are used. In physiology much attention has been paid to the methods by which data are secured, but the statistical evaluation is frequently neglected." . . . W. D. BATEN is now Chief, Operations Analysis Branch (A-5), Air Defense Command, Mitchell Field, New York. He was formerly with Michigan State College, East Lansing. . . . H. J. MILLER, Assistant Professor of Plant Pathology, The Pennsylvania State College, State College, is the only person (as far as we know) who reads the editorials in *Biometrics*. He states, "I wish to say that I have found this publication to be very useful in its present form. For the average biological worker it affords an opportunity to have questions answered by a competent statistician such as is not available in many of our institutions." Now, we will see if he reads the NEWS AND NOTES. . . . D. B. SHANK, who was Assistant Agronomist, University of Arkansas, Fayetteville, is now Associate Agronomist with the Agronomy Department, Agricultural Experiment Station, Brookings, South Dakota. His work consists of corn breeding and of corn variety testing. He finds different problems in South Dakota from those met in the South. For instance, "the short growing season and the resulting necessity for early hybrids is something that I did not have to worry about before." But look how level the land is out there, more homogeneous soil! . . . P. V. SUKHATME, statistical adviser, Imperial Council of Agricultural Research, New Delhi, India, reports that, "During the last few years the Statistical Section of the Imperial Council of Agricultural Research has carried out sample surveys for estimating the yield per acre of rice in all the rice-growing districts of the United Provinces, Central Provinces, Bombay, Madras, Orissa and Bihar." In the Annual Report (1946–1947) of the Social Science Research Center, University of Puerto Rico, Rio Piedras, of which CLARENCE SENIOR is Director, is found a report of a human biology study. HARRY SHAPIRO is directing a joint project with the Center, the American Museum of Natural History and the Department of Anthropology of Columbia University. The project is to examine the relationship between soil, water supply, food and the biological status of man, including the incidence of inherited defects. Anthropometric measurements will be made later in the study to check on nutritional inadequacies. . . . R. E COMSTOCK has returned to the Institute of Statistics as Professor of Genetic Statistics. For the last year and a half he has been at Agricultural Experiment Station, Rio Piedras, as Head of the Animal Industry Department. . . . M. S. BARTLETT has been appointed to a newly-established Professorship of Mathe-

matical Statistics at Manchester University. . . . The Statistical Summer Session at the Virginia Polytechnic Institute from August 5 to September 9 was attended by 166 persons from 33 states and 5 foreign countries. The list of courses included "Mathematics of the Design of Experiments" by R. C. BOSE of the University of Calcutta, India; "Design of Animal Experiments" by H. L. LUCAS, University of North Carolina, "Statistical Methods" by G. W. SNEDECOR of Iowa State College, "Mathematical Statistics" by G. W. BROWN of Iowa State, and "Engineering Statistics" by BOYD HARSHBARGER of Virginia Polytechnic Institute. There were three courses in Sampling covering the fields of Interview Technique, Sampling Methods, and the Mathematics of Sampling. They were given by Drs. METZNER and CANNELL of the University of Michigan and E. E. HOUSEMAN and W. A. HENDRICKS of the B.A.E. Among the seminar speakers were: W. F. CALLANDAR, MORRIS HANSEN, CHARLES F. SARLE, HAROLD HOTELLING, E. S. KEEPING of Canada, P. J. BULON, W. E. DEMING, G. W. SNEDECOR, G. W. BROWN, M. G. KENDALL, of England, R. C. BOSE of India, and GERTRUDE COX.

Officers of the American Statistical Association: President, Willard L. Thorp; Directors, Isador Lubin, Lowell J. Reed, Walter A. Shewhart, Samuel A. Stouffer, Helen M. Walker, Samuel S. Wilks; Vice-Presidents, Chester I. Bliss, Philip M. Hauser, Stacy May, Jacob Marschak, Jerzy Neyman, Frank W. Notestein, George W. Snedecor, Aryness Joy Wickens; Secretary-Treasurer, Lester S. Kellogg.

Officers of the Biometrics Section: Chairman, D. B. DeLury; Secretary, H. W. Norton; Section Committee, Geoffrey Beall, E. J. DeBeer, D. B. DeLury, D. J. Finney, H. W. Norton and J. W. Tukey.

Editorial Committee for *Biometrics:* Chairman, Gertrude M. Cox; Members, R. L. Anderson, C. I. Bliss, W. G. Cochran, Churchill Eisenhart, H. W. Norton, G. W. Snedecor and C. P. Winsor; collaborating editors, W. J. Dann, D. J. Finney, G. E. Dickerson, H. O. Halverson, C. M. Mottley, J. G. Osborne.

Material for *Biometrics* should be addressed to the Chairman of the Editorial Committee, Institute of Statistics, North Carolina State College, Raleigh, N. C.; and material for Queries should go to "Queries," Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.